



HAL
open science

Data littératie & SHS : développer des compétences pour l'analyse des données

Béa Arruabarrena, Gérald Kembellec, Ghislaine Chartron

► **To cite this version:**

Béa Arruabarrena, Gérald Kembellec, Ghislaine Chartron. Data littératie & SHS : développer des compétences pour l'analyse des données. CODATA, Mar 2019, Marne la Vallée, France. hal-03154432

HAL Id: hal-03154432

<https://hal.science/hal-03154432>

Submitted on 28 Feb 2021

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Data littératie & SHS : développer des compétences pour l'analyse des données

Conférence CODATA

DATA Value Chain in Science & Territories, Marne la vallée, 14-15 mars 2019

Béa Arruabarrena¹, Gérald Kembellec² and Ghislaine Chartron³

^{1,2,3}Laboratoire Dicen-IdF, Cnam, Paris, France

E-mail: prenom.nom@lecnam.net

Abstract

The main goal of this article is to analyze the skills required along the data value chain and to question them in epistemological and methodological ways to include them in a data analysts diploma. This new range of technical and methodological skills, called *data literacy*, is here put into relation with social analysis and, more generally, with the issues of the human sciences. Indeed, social and humanities aspects dealings with these subjects enrich the use of the data by methods of representation or knowledge organization, keeping in mind the legislative and ethical aspects that must lead the data strategy within an organization.

L'objectif principal de cet article est d'analyser les compétences requises tout au long de la chaîne de valeur des données et de les interroger, de manière épistémologique et méthodologique, pour les inclure dans un diplôme d'analyste de données. Cette nouvelle gamme de compétences techniques et méthodologiques, appelée « data littéracie », est ici mise en relation avec l'analyse sociale et, plus généralement, avec les enjeux des sciences humaines. En effet, ces derniers enrichissent l'utilisation des données par des méthodes de représentation ou d'organisation des connaissances tout en gardant à l'esprit les aspects législatifs et éthiques devant guider la stratégie de données au sein d'une organisation.

Keywords: Data literacy, data analyst, Big Data, quality of data, ethic.

Keywords (fr): Littératie numérique, analyste de données, méga données, qualité des données, éthique

Introduction

Avec l'avènement récent de l'ère dite du « Big Data », les données massives représentent un enjeu majeur pour la société. Le développement de nouveaux services émerge dans tous les secteurs. Que ce soient les entreprises, les organisations, ou encore les citoyens, tous sont devenus producteurs et consommateurs de données. La maîtrise des données n'est plus un problème exclusivement réservé aux sciences dites « dures », elle concerne désormais aussi les sciences sociales et humaines (Boyd et Crawford, 2012).

Avec les Mégadonnées, le traitement de données est devenu une compétence incontournable à acquérir pour les entreprises et les organisations dans tous les secteurs d'activités. L'émergence de nouveaux profils ayant des compétences pour produire, collecter, analyser, comprendre, et utiliser les données s'est affirmée. Le traitement de données repose sur un large panel méthodologique qui vise l'amélioration continue des services pour l'utilisateur en particulier l'analyse et le reporting décisionnels (Régnier-Pécastaing et al., 2008). Il s'agit d'un processus complexe qui nécessite le croisement de multiples compétences en lien avec l'informatique, les statistiques, mais également en lien avec une

vision stratégique des données de l'organisation. Cette stratégie se doit d'inclure la connaissance du cadre réglementaire (RGPD) et de l'éthique qui impacte la qualification des différentes sources de données, la maîtrise des méthodes de collecte et d'analyse de données.

1. Objectifs & méthodes

L'objectif de cette communication est d'analyser les compétences requises pour un Data analyst, et ce à tous les niveaux de la chaîne de valeur de la donnée (Miller, et al. 2013). Après un état de l'art sur le concept de Data littératie qui nous permet de cerner les enjeux liés à l'acculturation aux données, nous proposons un référentiel de blocs de compétences élaborés progressivement pour le Master « Mégadonnées et Analyse sociale » (MÉDAS)¹ depuis sa création en 2017, au CNAM-Paris. Dans une dernière partie, nous discuterons cette proposition à plusieurs niveaux : l'inclusion de compétences relevant des sciences humaines et sociales dans ce type de programme, le caractère transverse des compétences visées participant au renouveau de certaines d'entre elles, et les premiers éléments d'interaction entre la Data littératie, la littératie informationnelle et la littératie numérique.

Au plan méthodologique, c'est donc à l'appui de notre activité d'enseignement au sein du Master « Mégadonnées et Analyse sociale », que nous proposons une analyse des compétences de Data analyst. Le Master MÉDAS a pour vocation de former en apprentissage des étudiants venant à la fois des Sciences humaines et sociales, des Sciences informatiques et statistiques. L'objectif est de développer des compétences visant à répondre à des problématiques de données dans le contexte d'organisations publiques, privées ou associatives. Aujourd'hui, la demande pour ce type d'emploi est en plein essor aussi bien dans les secteurs de la banque, de l'assurance, de la santé, des médias, des organismes publics de statistiques, et des collectivités locales entre autres. Les débouchés au terme de la formation concernent en premier lieu, des postes de Responsable des données, de Manager des données (Chief Data officer), de Chargé d'études, ou encore de Data scientist.

2. Data littératie : état de l'art

Les travaux de recherches portant sur le concept de littératie de la donnée (Data littératie) sont nombreux et varient souvent en fonction du domaine étudié. Selon Mandinach et Gummer (2013), ce qui définit la Data littératie est « la capacité à comprendre et à utiliser les données de manière effective pour la prise de décision ». Cette définition se focalise volontairement sur l'usage et non sur les aspects techniques. Ce lien avec les Sciences de la décision est récurrent et traduit l'adossement de l'analyse des données à une vision stratégique, prolongeant ainsi les activités de veille informationnelle ou de prospective pour les organisations. Parmi les définitions en lien avec le travail d'analyse, certaines mettent l'accent en priorité sur les compétences mathématiques, informatiques et statistiques (Wolff et al., 2016). Néanmoins, d'autres auteurs montrent que les compétences de base d'une Data littératie ne peuvent être considérées de manière isolée (Matthew, 2016). Aussi, si ces compétences sont requises et qu'elles prennent une place considérable dans les programmes de formation, il apparaît également que des compétences transverses basées sur des connaissances issues des SHS, telles que l'épistémologie du Big data, le cadre juridique, le pilotage spécifique de projets (Shearer, 2000 ; Marr, 2015), la gouvernance, la qualité (Arruabarrena, 2018) et l'éthique des données (Matthew, 2016) sont indispensables à la maîtrise des données. Pour Calzada-Prado et Marzal (2013), le concept de Data littératie est la capacité des individus « d'accéder, d'interpréter, d'évaluer de manière critique, de gérer, de manipuler et d'utiliser éthiquement les données »². En ce sens, l'enjeu d'une data littératie est le développement « d'une culture des données », i.e. la capacité à évaluer de manière éthique et critique les données (Carlson et al., 2011).

¹ <http://formation.cnam.fr/rechercher-par-discipline/master-sciences-humaines-et-sociales-mention-humanites-numeriques-megadonnees-et-analyse-sociale-813168.kjsp>

² Traduction des auteurs

Sur cette question des enjeux de développement d'une Data littératie, Handel (2012) indiquent que la Data littératie doit s'accompagner de la maîtrise de l'information qui renvoie à une compétence transverse. Cette littératie informationnelle (Information literacy) doit être associée aux usages, au contexte - c'est-à-dire avec les connaissances du domaine étudié : les domaines sectoriels et métiers, ainsi que la stratégie de l'entreprise. Avant tout traitement et analyse, il faut être capable de formuler une question ou un problème reposant sur des compétences à appréhender et analyser un contexte, un type de gouvernance et les objectifs stratégiques de l'organisation (Anderson, et al. 2015, Nesvijevskaia, 2019), car c'est la définition préalable d'objectifs intelligibles aux analyses qui permettra la validation de ces mêmes objectifs (Tygel, et al., 2015). Certaines approches plus culturelles et sociales permettent justement de considérer que les apports de la technologie n'ont de sens que dans le cadre d'objectifs sociaux prédéfinis (Matthew 2016). Matthew explique que les compétences liées à la collaboration constituent une « compétence communautaire » comme socle de la compréhension et l'interprétation des données. Qui plus est, la dissociation des données de leur contexte peut rendre l'analyse difficile, et la réutilisation impossible (Schneider, 2013) avec le risque d'être mal interprétées (Bradshaw, 2014). Cela nécessite que la conduite d'un projet prenne en compte dans sa méthodologie à la fois les points de vue des producteurs et des usagers de données (Carlson, et al. 2011).

Cette littératie intervient dans la compréhension des données pour guider le choix des analyses et de leur modalité de restitution. Cela signifie qu'il faut non seulement comprendre et analyser données dans leur contexte informationnel de production et d'usages, mais aussi être capable d'en restituer les résultats et de communiquer sur plusieurs supports (présentations, rapport, visualisation).

Parmi les tentatives de définitions d'une littératie des données massives, la littératie statistique a souvent été sous-estimée, laissant croire que la quantité pouvait suffire à apporter des réponses aux différentes questions posées par la société, préférant se concentrer sur l'exploration et la découverte fortuite de relations entre données (Carlson et al., 2011). Dans de nombreux domaines (Médecine, Pharmacologie, Biologie, Chimie, etc.) les méthodes de Datamining et de Machine Learning sont performantes. Cependant, de nombreux cas reposent essentiellement sur des méthodes statistiques, notamment des approches analytiques et décisionnelles, celles d'enquêtes sociales, ou encore celles de prétraitement (échantillonnage, statistiques descriptives, etc.). Ces deux approches sont néanmoins souvent complémentaires. Ce constat nous renvoie à un aspect essentiel de la maîtrise des données : une bonne compréhension du phénomène du Big Data, en particulier le fait que la quantité de données n'est pas forcément un gage de fiabilité pour les analyses (Koltay, 2015). Au contraire, l'étude des pratiques de scoring bancaire aux États-Unis (Pasquale, 2015) montre que les Mégadonnées peuvent dans certains cas être à la source de « méga-erreurs » (Liu et al., 2016). Ces inexactitudes peuvent être dues au mode de sélection des données, à la disparité de leur qualité ou encore à l'usage d'algorithmes, souvent porteurs d'implicites économiques, sociaux et culturels qui orientent l'analyse et parfois l'altèrent (Cardon, 2015, Abitboule, 2017). Ces biais ayant des conséquences éthiques directes sur l'interprétation des données, il est donc important de bien sélectionner et comprendre les algorithmes et d'être à même de questionner la qualité des données pour évaluer la confiance à leur accorder.

Une autre compétence importante est celle liée à la communication et la diffusion des résultats d'analyse en particulier grâce aux nombreux outils de visualisation et d'exploration de données disponibles (R, Python, Tableau, MS-Power BI, etc.). Il existe une tendance au design d'information et au Storytelling qui vise à produire une narration avec les données. Cette approche est considérée comme le moyen le plus efficace de représenter le cycle de vie de la donnée dans son contexte. Il est ainsi possible de représenter le traitement permettant de répondre aux questions posées initialement – l'objectif initial – mais aussi de proposer de nouvelles perspectives mises en lumière par l'analyse

(IODC, 2015)³. La réussite de la création d'une visualisation des données efficace repose essentiellement sur la compréhension des principes fondamentaux de la sémiologie graphique (Bertin, 1967). En effet, les différents choix de sélection et de présentation des données auront un impact fort sur leur réception. Ainsi, la texture, la couleur, les proportions d'un graphe vont possiblement orienter l'analyse des décideurs. Ces choix communicationnels de rendus visuels ont donc un impact stratégique et font parfois appel à l'éthique de l'analyste tout autant qu'à sa capacité à communiquer ses analyses.

3. Focus sur 5 compétences de référence pour le *Data analyste*

Le référentiel du Master MÉDAS repose sur un découpage en cinq blocs de compétences qui correspondent à la fois à des compétences intervenant dans la chaîne de valeur du traitement des données, mais également à des compétences transverses à de nombreux métiers impactés aujourd'hui par les données.

Compétence n°1 : Collecter et préparer des données en vue d'un traitement informatisé

La collecte est une étape du traitement des données cruciale et déterminante pour toute la chaîne de valeur, les données étant plus souvent "captées" que "données" selon l'idée du *capta (vs) data* de Drücker : « *Capta is "taken" actively while data is assumed to be a "given" able to be recorded and observed* » (Drücker, 2011) dans laquelle les données doivent être patiemment "obtenues", ce qui signifie un important travail en amont (Latour, 2012). La pertinence des résultats des analyses de données dépend donc directement de la qualité des données collectées. Cet écueil du traitement de données est connu, trop de données sont collectées sans pouvoir être exploitées, car de qualité insuffisante. La qualité des données est donc également un concept central pour la maîtrise d'une chaîne de traitement analytique (Koltay, 2016). Cette étape sous-entend une compétence souvent sous-estimée alors qu'elle constitue une partie considérable du travail sur les données, préalable et nécessaire à toute analyse (Coulton, et al., 2015 ; Kingsley, et al., 2013). Elle nécessite également des compétences sur les méthodes de production, les technologies spécifiques d'extraction (ETL), de scraping⁴ et d'API⁵ par exemple. Le nettoyage et la préparation des données permettent de les transformer dans des formats adaptés pour l'import par les dispositifs d'analyses. Il s'agit également, au regard des SHS, d'avoir une connaissance fine des données et de prendre autant en compte le point de vue de l'utilisateur que de celui du producteur de données.

Les principales compétences de la collecte de données concernent les actions suivantes :

1. Requête des bases de données et en exporter les données pour constituer son corpus ;
2. Scraping : développer des scripts spécifiques (python par exemple) ;
3. S'approprier les méthodes de modélisations de données structurées (Merise, UML) et non structurées (NoSQL), ainsi que leurs encodages (utf-8, iso-8859-1...) et formats (SQL, JSON, CSV, XML...) afin de pouvoir les extraire et les réexploiter ;
4. Nettoyer/Préparer
 - Normaliser les données afin de pouvoir traiter des ensembles de données cohérents
 - Encoder, transcoder des fichiers dans divers formats de stockage (Excel, CSV, JSON)
 - Réconcilier les données en regroupant des jeux issus de différentes sources et en les enrichissant à partir de référentiels d'autorités.

Compétence n°2 : Analyse des données pour produire des informations, des indicateurs utiles aux organisations ou à des individus

³ <http://1a9vrva76sx19qtvgl1ddvt6f.wpengine.netdna-cdn.com/wp-content/uploads/2015/11/opendatacon-report-en-web.pdf> (consulté le 12/2/2019)

⁴ Scraping : technique de collecte et d'extraction de données dans des corpus bruts, comme les pages Web.

⁵ API (*Application Programming Interface*) : service web pensé par et pour les développeurs afin d'offrir à un programme client un point d'entrée accédant à des données filtrées et structurées.

Du point de vue des SHS, l'enjeu principal pour l'analyse de données est de répondre à des questions de société à partir de données quantitatives et/ou qualitatives (base de données, statistiques...). Il s'agit donc de concevoir et d'évaluer de manière critique des analyses de données. Deux types d'enseignement statistiques et datamining cohabitent aujourd'hui dans les sciences de données :

1. Analyser des données via des méthodes statistiques dans une démarche descriptive ou analytique (issues de la business intelligence (BI), statistiques descriptives, inférentielles, etc.) afin d'affirmer ou d'infirmer des hypothèses et/ou stratégies en lien avec les besoins des métiers composant une organisation (SAS, Excel, logiciel R, etc.). Les méthodes statistiques cherchant à répondre à une question posée a priori ou à valider une hypothèse dans un cadre défini (ex. échantillonnage pour la représentativité);
2. Analyser des données via des méthodes de Datamining, de Machine Learning et d'Intelligence artificielle dans une démarche exploratoire (logiciel R, Python) : ces méthodes fonctionnent sur l'exploration et la découverte de relations entre données ou de modèles, de manière heuristique sans hypothèses forcément définies au préalable.
 - Analyser à l'aide des techniques de Datamining (fouille de données, analyse spatiale et temporelle, analyse de réseaux, cartographie) afin de faire émerger de nouvelles connaissances ou de nouveaux usages sociaux ;
 - Sélectionner et utiliser ou éventuellement implémenter des algorithmes de traitements des données à l'aide des méthodes d'apprentissage automatique (Machine Learning, Deep Learning, réseaux de neurones).

Ces méthodes, bien que partant d'un postulat différent, sont souvent amenées à se combiner et se compléter dans un projet d'analyse de données. Elles s'appuient à la fois sur la programmation de type Python, les logiciels statistiques comme R. On note un certain engouement dans le domaine de l'analytcs/BI pour des logiciels et plateformes d'analyse d'aide à la décision qui sont de plus en plus accessibles et performants (Tableau, MS Power BI, etc.). Néanmoins, ces derniers, même s'ils permettent un très bon niveau d'analyse, ne peuvent être considérés comme la seule facette des compétences d'un Data analyst. Il est important de préciser qu'une compréhension fonctionnelle des algorithmes (classification, régression, etc.) et de leurs applications est incontournable (Description, Prédiction, Prescription). Il est également nécessaire de connaître les usages les plus répandus en analyse de données, tels que les analyses prédictives de l'attrition par exemple⁶.

Compétence n°3 : Pilotage de la qualité des données avant et après le traitement des données au regard des valeurs sociétales

Comme explicité précédemment, la qualité des données est essentielle à leur analyse, aussi bien du point de vue technique (Berti-Equille, 2004) que de celui de l'analyse sociale (Arruabarrena, 2018). Aussi, en quelques années la gouvernance et la qualité des données sont devenues des thèmes centraux tant pour les activités professionnelles que scientifiques. Cela démontre l'intérêt au plan stratégique d'établir un référentiel commun pour garantir que les données soient disponibles et de bonne qualité, y compris au sens éthique du terme.

Les principales compétences pour le pilotage de la qualité des données sont :

1. Développer une analyse critique et éthique sur les données afin de constituer des corpus qui seront validés par les organisations
2. Sélectionner les données en respectant la législation en vigueur (RGDP), pouvant être diffusées en interne ou en externe dans l'organisation afin d'assurer un traitement des données en accord avec la politique et la stratégie de l'entreprise/organisme.
3. Gérer les données (Master Data Management)
 - Concevoir ou sélectionner ou des référentiels à l'aide de briques logicielles de type Talend
 - Décrire les données et intégrer des métadonnées
4. Établir un plan de gestion des données et/ou des livrables documentant les données et les processus associés (Data-Book)

⁶ L'attrition désigne l'analyse de la perte de clients ou d'abonnés.

Compétence n°4 : Gestion d'un projet « data » dans une organisation

Cette compétence transversale est indispensable pour assurer la conduite d'un projet dans le cadre de la stratégie data de l'entreprise. Si les méthodes de management de projets classiques et agiles sont utiles, il faut également signaler les méthodes d'émergence et de priorisation de portefeuille de projets. Il s'agit de gérer des projets de données articulés en fonction des besoins métiers. Des méthodes plus spécifiques de gestion de « projet-données » sont également mises en œuvre (Marr, 2015, Shearer, 2000), ou en cours de définition (Nesvijevskaia, 2019).

Les principales compétences pour mener un projet data concernent les étapes suivantes :

1. Diagnostiquer, auditer, repenser la structuration de la data de l'entreprise
2. Dialoguer avec les clients/métiers et analyser leurs besoins et les usages associés
3. Assurer la communication en présentant les données sous différentes formes visuelles (tableaux, graphiques, ...)
4. Interpréter les résultats et rédiger un livrable de synthèse préconisant des solutions d'aide à la décision et contribuer à l'optimisation de la stratégie de l'organisation
5. Identifier et développer de nouveaux leviers de croissance associés à l'usage de données.

Compétence n°5 : Conception des services de données pour transformer des activités sociétales

Cette compétence est tournée vers l'usage des données et l'innovation de services. Il s'agit à partir de l'expression des besoins de proposer, de modéliser et de concevoir des services data spécifiques à l'entreprise et à son activité. Plus globalement, il s'agit de développer des stratégies data dans des contextes variés en tenant compte des contraintes opérationnelles (financières, humaines, économiques, sociales...). La conception d'interfaces et de datavisualisations joue un rôle important dans la diffusion des analyses et l'usage de la donnée. Les principales compétences pour la conception de services sont :

1. Identifier la valeur des données dans différents contextes d'activités, et transformer les activités
2. Identifier le périmètre du service (cerner ses limites) et évaluer sa capacité à résoudre des problématiques posées par des individus ou des collectifs
3. Développer des interfaces Hommes-Données : Élaborer des tableaux de bord et diffuser les données (et rapports) en restituant les résultats par des techniques de visualisation.

4. Conclusion

Notre analyse de la littérature scientifique montre que si les compétences constitutives d'une Data littératie peuvent être techniques, informatiques et statistiques (compétences 1 et 2), et qu'elles constituent une dimension importante des programmes de formation des Data analysts, elles ne suffisent pas à développer des analyses pertinentes pour la société. En effet, des compétences en SHS sont indispensables à la compréhension des données et l'appréhension de leurs usages pour la société. Il s'agit non seulement d'acquérir des compétences de niveau épistémologique, éthique, et juridique, mais également des compétences de pilotage de la qualité des données (compétence n°3), de méthodes de gestion de projets « data » (compétence n°4) et de conception de services de données (compétence n°5). Aussi, le caractère transversal des compétences visées participe au renouveau de certaines d'entre elles, comme le montre la combinaison des méthodes statistiques avec celles du datamining.

Dans la continuité de notre raisonnement sur la nécessité d'inclure plus en avant les SHS dans les programmes de formation des Data analysts, et comme le souligne un certain nombre d'auteurs, la maîtrise des données est directement liée à une maîtrise de l'information du domaine analysé (i.e. au développement d'une littératie informationnelle). Haendel et al. (2012) qui s'intéressent à la Data littératie dans le domaine de la recherche, parlent à ce propos d'une « culture scientifique sémantique », constitutive de sens commun, et affirment qu'une littératie de données et une littératie

informationnelle devraient accompagner la formation scientifique afin d'établir une nouvelle norme culturelle socle indispensable pour la compréhension et l'interprétation des données. On voit donc se dessiner une data littératie basée sur le développement d'interactions étroites entre littératies informationnelle et numérique. Ces interactions sont aux fondements d'une réflexivité critique que sous-tend cette nouvelle forme de littératie (Buschman, 2009).

Notre analyse vise à se poursuivre, en effet cette proposition de référentiel est en cours d'expérimentation depuis 3 ans au sein du Master MÉDAS. L'approche pragmatique que nous avons engagée permet des réajustements progressifs dans un contexte en évolution très rapide. Nous envisageons notamment de mettre l'accent sur les enjeux et les méthodes de la sociologie d'enquête ainsi que sur l'éthique et la philosophie de la donnée.

References

- Abiteboul, S., Doweck, G. (2017), *Le temps des algorithmes*, Paris, Le pommier.
- Anderson, R., Gummer, E. S., Mandinach, E. B., & Parton, B. M. (2015). *Ethical and Appropriate Data Use Requires Data literacy*. *Phi Delta Kappan*, Vol. 96 n°5, pp. 25–28.
- Arruabarrena (2018), Big data & éthique : la qualité des données en débat. In *L'Éthique en contexte info-communicationnel numérique* (Dir L. Balicco, E. Broudoux, G. Chartron, V. Clavier, I. Paillart), De Boeck, pp. 25-18.
- Berti-Equille, L.(2004), Un état de l'art sur la qualité des données. *Ingénierie des systèmes d'information*, Vol. 9 n°5 et 6, pp. 117-143.
- Bertin, J., *Sémiologie graphique*, Paris, Mouton/Gauthier-Villars, 1967.
- Boyd, D., Crawford, K. (2012), Critical questions for big data: provocations for a cultural, technological, and scholarly phenomenon, *Information, Communication & Society*, Vol. 15 n°5, pp. 662-679.
- Bradshaw, L. (2014). Beyond Data Science: Advancing Data literacy – Made by Many – Medium. consulté en ligne le 12/2/2019 <https://medium.com/the-many/moving-from-data-science-to-data-literacy-a2f181ba4167#.ayhkrcoop>
- Buschman, J. (2009), Information literacy, new literacies, and literacy, *Library Quarterly*, Vol. 79 No. 1, pp. 95-118.
- Calzada Prado, J. and Marzal, M.A. (2013), Incorporating Data literacy into information literacy programs: core competencies and contents, *Libri*, Vol. 63 n°2, pp. 123-134.
- Cardon, D.(2015), *A quoi rêvent les algorithmes. Nos vies à l'heure : Nos vies à l'heure des Big data*. Le Seuil, Paris.
- Carlson, J., Fosmire, M., Miller, C. C., & Nelson, M. S. (2011). Determining Data Information Literacy Needs: A Study of Students and Research Faculty. *Portal: Libraries and the Academy*, Vol. 11, n°2, 629–657.
- Coulton, C. J., Goerge, R., Putnam-hornstein, E., & de Haan, B. (2015). Harnessing Big Data for Social Good : A Grand Challenge for Social Work, *American Academy of Social Work and Social Welfare*, pp. 1-20, Cleveland.
- Dalbello, M. (2011), A genealogy of digital humanities, *Journal of Documentation*, Vol. 67 No. 3, pp. 480-506.
- Davies, T., Perini, F., & Alonso, J. (2016). Researching the emerging impacts of open data ODDC conceptual framework. *The Journal of Community Informatics*, 12, 1–37.
- Drucker J. (2011). Humanities approaches to graphical display. *Digital Humanities Quarterly*, Vol. 5, n°1, pp. 1–21.
- Haendel M., Vasilevsky N., Wirz J. (2012), Dealing with Data: A Case Study on Information and Data Management Literacy. *PLoS Biol* Vol. 10, n°5 Consulté en ligne le 12/2/2019. <https://doi.org/10.1371/journal.pbio.1001339>
- Heinderyckx (F.), *Le tournant numérique*, *Hermès*, vol. 71, n°1, pp. 87-91, 2015.
- Hunt, K. (2004), The challenges of integrating Data literacy into the curriculum in an undergraduate institution, *IASSIST Quarterly*, Vol. 28, n°2, pp. 12-15.
- IOCD - International Development Research Centre. (2015). *Enabling the Data Revolution : An International Open Data Roadmap*, in (éd. Tim Davies) Report of the 3rd conference on Open Data, Ottawa, Canada. Consulté en ligne le 12/2/2019 <http://1a9vrva76sx19qvtvg1ddvt6f.wpengine.netdna-cdn.com/wp-content/uploads/2015/11/opendatacon-report-en-web.pdf>
- Kingsley, G. T., Pettit, K. L., & Hendey, L. (2013). *Strengthening Local Capacity For Data-Driven Decisionmaking*, rapport de The Urban Institute, consulté en ligne le 12/2/2019: https://www.urban.org/research/publication/strengthening-local-capacity-data-driven-decisionmaking/view/full_report
- Koltay, T. (2016). Data governance, Data literacy and the management of data quality. *IFLA Journal*, Vol. 42, n°4, pp. 303-312.

- Latour B., Jensen P., Venturini T., Grauwin S., Boullier D. (2012), The Whole is Always Smaller Than Its Parts : How Digital Navigation May Modify Social Theory. (with Pablo Jensen, Tommaso Venturini, Sébastien Grauwin and Dominique Boullier), *British Journal of Sociology*, Vol. 63 n°4.
- Liu, J., Li, J., Li, W., & Wu, J. (2016). Rethinking big data: A review on the data quality and usage issues. *ISPRS Journal of Photogrammetry and Remote Sensing*, n° 115, pp. 134-142.
- Mandinach, E. B., & Gummer, E. S. (2013). A Systemic View of Implementing Data literacy in Educator Preparation. *Educational Researcher*, Vol. 42, n°1, pp. 30–37. <http://doi.org/10.3102/0013189X12459803>
- Marr, Bernard (2015). *Big Data: Using SMART Big Data, Analytics and Metrics To Make Better Decisions and Improve Performance*. John Wiley & Sons.
- Matthews, P. (2016) Data literacy conceptions, community capabilities. *The Journal of Community Informatics*, Vol. 12, n°3.
- Miller, H. G., et P. Mork. 2013. « From Data to Decisions: A Value Chain for Big Data ». *IT Professional*, Vol. 15, n°1, pp. 57-59.
- Nesvijevskaia, A. (2019) *Phénomène Big Data en entreprise : processus projet, médiations et impact sur les indicateurs de génération de valeur* (thèse CNAM à paraître, s.d. G. Chartron),
- Pasquale, F.(2015), *The black box society: The secret algorithms that control money and information*. Harvard University Press.
- Qin, J. and D'Ignazio, J. (2010), "Lessons learned from a two-year experience in science Data literacy education", Proceedings of the 31st Annual IATUL Conference, 20-24 June, consulté en ligne le 12/2/2019: <http://docs.lib.purdue.edu/iatul2010/conf/day2/5>
- Régnier-Pécastaing, F., Gabassi, M., et Finet, J.,(2008), *MDM: Enjeux et méthodes de la gestion des données*. Dunod.
- Schafer, V., & Thierry, B. (2015). L'ogre et la toile. Le rendez-vous de l'histoire et des archives du web. *Socio. La nouvelle revue des sciences sociales*, n°4, pp.75-95.
- Shearer, Colin. 2000. The CRISP-DM model : the new blueprint for data mining. *Journal of Data Warehousing*, Fall 2000, A 101 communications Publication edition, sect. Vol. 5, n°4.
- Schneider, R. (2013), Research Data literacy, in Kurbanoglu, S. et al. (Ed.), *Worldwide Commonalities and Challenges in Information Literacy Research and Practice*, Springer International, Cham, pp. 134-140.
- Tygel, A. F., Campos, M. L. M., & de Alvear, C. A. S. (2015). Teaching Open Data for Social Movements: a Research Strategy. *The Journal of Community Informatics*, Vol. 11, n°3.
- Wolff, Annika & Moore, John & Zdráhal, Zdenek & Hlosta, Martin & Kuzilek, Jakub. (2016). *Data literacy for learning analytics*, pp. 500-501. 10.1145/2883851.2883864.